

**Modeling Motive Activation in the Operant Motives Test: A Psychometric Analysis Using
Dynamic Thurstonian Item Response Theory**

Accepted in Motivation Science.

J. Malte Runge and Jonas W. B. Lang

Ghent University

Stefan Engeser

University of Jena

Julia Schöler

University of Konstanz

Sophie C. den Hartog

Ghent University

Ingo Zettler

University of Copenhagen

Author Note

J. Malte Runge, Jonas W. B. Lang and Sophie C. den Hartog, Department of Personnel
Management, Work and Organizational Psychology, Ghent University; We thank Christian

Ewen, Christian Helm, Johannes Kleikamp, and David Kliegel for their help in coding and collecting data.

Correspondence concerning this article should be addressed to Malte Runge, Department of Personnel Management, Work and Organizational Psychology, Ghent University, 9000 Ghent

Contact: malte.runge@ugent.be

Abstract

The Operant Motive Test (OMT) is a picture-based procedure that asks respondents to generate imaginative verbal behavior that is later coded for the presence of affiliation, power, and achievement-related motive content by trained coders. The OMT uses a larger number of pictures and asks respondents to provide more brief answers than earlier and more traditional picture-based implicit motive measures and has therefore become a frequently used measurement instrument in both research and practice. This article focuses on the psychometric response mechanism in the OMT and builds on recent advancements in the psychometric modeling of the response process in implicit motive measures through the use of Thurstonian item-response theory. The contribution of the article is twofold. First, the article builds on a recently developed dynamic Thurstonian model for more traditional implicit motive measures (Lang, 2014) and reports the first analysis of which we are aware that applies this model to OMT data ($N = 633$) and studies dynamic motive activation in the OMT. Results of this analysis yielded evidence for dynamic motive activation in the OMT and showed that simulated IRT reliabilities based on the dynamic model were .52, .62, and .73 for the affiliation, achievement, and power motive in the OMT, respectively. The second contribution of this article is a tutorial and R code that allows researchers to directly apply the dynamic Thurstonian IRT model to their data. The future use of the OMT in research and potential ways to improve the OMT are discussed.

Keywords: Operant Motive Test, implicit motives, Thurstonian choice model, psychometric theory, Item Response Theory

Modeling Motive Activation in the Operant Motives Test: A Psychometric Analysis Using Dynamic Thurstonian Item Response Theory

Motivational researchers have long been interested in implicit motives that are typically described as stable affective preferences for certain goal states (McClelland, Koestner, & Weinberger, 1989; Schultheiss, Liening, & Schad, 2008; Winter, John, Stewart, Klohn, & Duncan, 1998). Implicit motives predict performance in the laboratory (e.g., McAdams, Jackson, & Kirshnit, 1984), psychological well-being (e.g., Baumann, Kaschel, & Kuhl, 2005), and are useful in the prediction of outcome criteria like job performance (e.g., Lang, Zettler, Ewen, & Hülshager, 2012), entrepreneurial success (e.g., Spangler, 1992), career success (McClelland & Boyatzis, 1982) or satisfaction in relationships (e.g., Winter et al., 1998). Implicit motives are also related to hormonal and neuronal indicators (Hall, Stanton, & Schultheiss, 2010; Kuhl & Kazén, 2008).

Most implicit motive measures are picture-based procedures (McClelland et al., 1989; Winter et al., 1998). These procedures consist of a series of pictures showing persons in different situations and ask respondents to describe what they see in the picture, what happened before, and what will happen next. The stories that respondents generate are then evaluated for their motive content by trained coders using standardized and detailed scoring manuals. Coders typically show good inter-rater agreement (Köllner & Schultheiss, 2014; Schüler, Brandstätter, Wegner, & Baumann, 2015).

The most common picture-based measure to assess implicit motives is the picture story exercise (PSE). The PSE was developed in the 1950s (McClelland, Atkinson, Clark, & Lowell, 1953), consists of a total of six pictures, and asks respondents to write elaborate stories in response to each of the pictures. A more recently developed implicit motive measure is the

Operant Motive Test (OMT). The OMT uses the same picture-based measurement approach like more traditional PSE-measures but uses a larger number (15) of simplified pictures and asks respondents to write only brief responses to these pictures. These characteristics may have contributed to the fact that the OMT is currently frequently used in implicit motive research (e.g., Baumann et al., 2005; Baumann & Scheffer, 2010; Hofer, Busch, Chasiotis, Kärtner, & Campos, 2008; Kazén & Kuhl, 2005; Lang et al., 2012; Schüler, Job, Fröhlich, & Brandstätter, 2008). The larger number of pictures in the OMT makes this measure also especially interesting for psychometric research.

One major problem of picture-based procedures like the PSE and the OMT has long been that the psychometric process underlying the responses in implicit motive measures is complex, and, in turn, has been difficult to capture (e.g., Entwisle, 1972). However, researchers have recently built on earlier theoretical work suggesting that people choose among different motivational alternatives that are not independent (Atkinson & Birch, 1970), and have applied Thurstonian item-response models (Thurstonian IRT) to model this process (Lang et al., 2012). More recently, research (Lang, 2014) has developed a dynamic Thurstonian IRT model for the more traditional PSE. This dynamic Thurstonian model additionally includes a dynamic IRT process (Verguts & De Boeck, 2000; Verhelst & Glas, 1993) and is inspired by the idea that motivational tendencies are dynamic and that their strength declines when they are acted out (Atkinson & Birch, 1970; Tuerlinckx, De Boeck, & Lens, 2002).

The first major contribution of this paper is to apply the dynamic Thurstonian IRT model (Lang, 2014) developed for the PSE also to the OMT. The OMT is now a frequently used measure and we accordingly believe that motivational research would benefit from additional psychometric insights into this measure. A second contribution of this paper is an accessible

description and demonstration on how researchers can apply the dynamic Thurstonian IRT model to their OMT data. This demonstration should help motivational researchers in using the dynamic Thurstonian IRT model in their own research.

Implicit Motive Measures

Implicit motives are described as people's wishes and desires (Winter et al., 1998) or "the disposition to be concerned with and to strive for a certain class of incentives or goals" (Emmons, 1989). The dispositions are implicit because they are usually not accessible through introspection (McClelland et al., 1989; Winter et al., 1998), and thus need to be inferred indirectly from coding imaginative verbal material. Implicit motive measures typically ask respondents to write a fantasy story based on ambiguous picture stimuli. These stories are then evaluated for motive content. Most research focuses on three basic motive contents: *need for affiliation*, the desire to build, maintain, and deepen social relationships; *need for achievement*, the desire to improve one's performance; *need for power*, the desire to influence other people (Schultheiss, 2008). For sake of simplicity, we refer to the three basic implicit motives as affiliation, achievement and power and omit the term "need for" in the remainder of this article.

Implicit motive measures are typically not or only weakly correlated with explicit questionnaire measures of motives (Köllner & Schultheiss, 2014; Spangler, 1992). Several authors have suggested that this finding is not surprising because implicit motive measures and questionnaires tap different types of individual differences (McClelland et al., 1989; Winter et al., 1998) that predict different classes of behavior.

Morgan and Murray's (1935) Thematic Apperception Test (TAT) was the first measure to use picture stimuli in order to measure subconscious aspects of personality and influenced future measures of implicit motives. The first adaption developed to measure implicit achievement,

affiliation, and power motivation is McClelland's Picture-Story Exercise (McClelland, 1987; McClelland et al., 1953). A more recent implicit motive measure is the OMT (Kuhl & Scheffer, 2002).

The Operant Motive Test

The OMT uses a larger number of pictures (typically 15 pictures instead of 6 in the PSE) that depict scenes with one or several persons (Kuhl & Scheffer, 2002; Scheffer, 2005). Participants are instructed to choose one character shown in the picture as their protagonist. In contrast to other implicit motive measures that ask for full and long explanations like the PSE, the OMT asks for brief answers to the questions: (1) "What is important for the person in this situation and what is the person doing?" (2) "How does the person feel?" (3) "Why does the person feel this way?" (4) "How does the story end?". The OMT developers chose this procedure in the hope to get a more spontaneous answer from participants, which is supposedly linked more directly to implicit motives (Baumann, Kaschel, & Kuhl, 2010). Another unique characteristic of the OMT is that the pictures are drawings instead of photographs so that characters are depicted in a neutral manner in order to facilitate respondent's identification with the chosen character. Figure 1 shows a sample picture of the OMT.

The respondents' answers are categorized into one of the following motivational themes: affiliation, achievement and power. Only one motive response is coded for each picture. Stories with a lack of these motivational themes are categorized as zero. Additionally, the OMT manual differentiates between motive implementation strategies via three approach, one avoidance, and one in-between category within each motive (Kuhl & Scheffer, 1999, 2002). Researchers frequently combine the approach categories in order to get an overall approach factor (Baumann et al., 2005; Kazén & Kuhl, 2011; Lang et al., 2012; Schüler et al., 2008; Schüler, Job, Fröhlich,

& Brandstätter, 2009). The approach codings for each picture are then added up to a *sum score* for each motive. The avoidance categories are usually analyzed separately.

Reliability of Implicit Motive Measures

The most frequent method to estimate reliability is internal consistency (Cronbach, 1951), which several researchers (Entwisle, 1972; Fineman, 1977; Lilienfeld, Wood, & Garb, 2000) found to be low—in the .20 to .40s—for implicit motive measures. However, internal consistency only equals the reliability when the underlying disposition is unidimensional. Internal consistency cannot deal with a scenario in which motive reactions depend on the activation of other motives and in which individuals react to motivational stimuli.

Modeling Motive Activation in the OMT Using IRT

Thurstonian IRT Models

Researchers have long recognized that respondents typically write stories that focus on one motive-related theme (e.g., either achievement or power). Motives may thus compete for activation in a particular story and respondents effectively choose between different motives when they respond to implicit motive measures (Atkinson & Birch, 1970; Atkinson, Bongort, & Price, 1977). An established tool for modeling complex choice behavior are Thurstonian models (e.g., Böckenholt, 2006; Brown & Maydeu-Olivares, 2011) and researchers have recognized that Thurstonian models can also be used to model choice behavior in implicit motive measures (Lang, 2014; Lang et al., 2012). Thurstonian models consist of pairwise comparisons and place these comparisons on a utility scale (Thurstone, 1927). This basic modeling approach can be used to also model more complex choice behavior among various response options. For instance, let a , b , c , and d be the options among which a person can choose. When a person is asked to bring a , b , c , and d into a rank order, this complex choice behavior can be recoded into a total of

six pairwise comparisons (does the person prefer a over b , a over c , a over d , b over c , b over d , and c over d). Full rank information is not always available.

However, Thurstonian models can also be used when not all pairwise comparisons are available due to design. For instance in the OMT, let a , b , c , and d be the options in a picture i . The answer a in picture i informs about three of six possible comparisons: the option a in picture i is stronger than option b , c , and d . The information that these pairwise comparisons yield is that these options were stronger than the respective other options.

The original Thurstonian model was a model that was designed to study the rank order of choices within a population. For instance, Thurstone was interested in people's general preferences for specific stimuli in apperception (Thurstone, 1927). Standard Thurstonian models thus do not take individual differences into account. Modern applications of Thurstonian models recognized that individuals can also differ in their preference for a particular option and thus transform Thurstonian models into Thurstonian IRT (Böckenholt, 2004, 2006, A. Brown & Maydeu-Olivares, 2011, 2013).

Applying the Thurstonian IRT model to the OMT

In this section we describe how OMT data can be analyzed using the dynamic Thurstonian IRT model for implicit motive measures that has earlier been described in the literature and applied to data for the PSE (Lang, 2014). To fully analyze data for an implicit motive measure using this model, it is frequently useful to test a series of nested models that includes more simple models. The most basic Thurstonian IRT model (M_1) includes a level 1 specification similar to classic Thurstonian models that do not differentiate between persons (Critchlow & Fligner, 1991). This model can be extended by adding level 2 specifications to include individual differences between persons (M_1 ; Böckenholt, 2001, 2006) and picture effects

(M_2). The basic models (M_1 and M_2) provide the basis for specifying and testing dynamic effects in the Thurstonian IRT Model (Lang, 2014). We test two variants of dynamic effects in a replication of previous research (Lang, 2014). In the first variant, the dynamic effect is *temporary* and only lasts for one Picture (M_3), in the second, variant, the dynamic effect is a *sustained* effect that adds up over the complete test length (M_4). The original empirical analysis comparing these different conceptualizations of dynamic effects found that the latter (the sustained dynamic effect) described PSE data better than the former. Below, we provide additional detail on these models. We start with a non-technical description and follow up with more statistical details. Available estimation methods are described in the demonstration below.

Level 1. The basis for applying Thurstonian IRT to the OMT and other implicit motive measures, is the insight that a motive response to a particular picture does not only provide information about the motive that is chosen in a given picture but also about the magnitude of the other motives. For instance, when an answer from person s in picture i is coded as affiliation, one gets information about comparisons between three categories: The preference for affiliation in option i is higher than preference for achievement, power, and other content. This reflects the idea of motive competition and is represented in level 1 of the generalized linear mixed-effects model.

Thurstonian IRT Model: M_1 . In order to account for individual differences in preferring one motive category (*aff* for affiliation, *ach* for achievement, *pow* for power) to another this model needs a person parameter, also known as θ in IRT. The person parameter in M_1 makes the Thurstonian model an IRT model (Boeck, 2008; Doran, Bates, Bliese, & Dowling, 2007).

Thurstonian IRT Model: M_2 . A main feature of IRT models is the analysis of picture effects, which are studied as cue strength in implicit motive research. To account for differences in mean evaluations of the motivational choice behavior options between pictures, a picture covariate is added in M_2 .

Dynamic Thurstonian IRT Models: M_3 and M_4 . Motivational researchers have long suggested that the strength of motives changes over time (Atkinson & Birch, 1970; Atkinson et al., 1977; Murray, 1938). One core idea has been the notion of dynamic effects. When a motive is acted out, the motive temporally loses some of its strength and behavior motivated by the respective motive becomes less likely. Behavior motivated by other motives becomes then more likely because the motives compete to be acted out.¹ After a refractory period, the strength of a motive may grow back to its previous strength. On the basis of Atkinson's theory, researchers have developed dynamic IRT models for implicit motive measures (Tuerlinckx et al., 2003, Lang, 2014). The work on dynamic Thurstonian IRT models on which we build (Lang, 2014) includes two different dynamic Thurstonian IRT models that differ in the length of their refractory period.

In the first dynamic Thurstonian model, M_3 , the dynamic effect is *temporary* and lasts only for one picture – motives recover quickly. We consider only the motive response from the last picture for this temporary dynamic effect.

In the second dynamic Thurstonian model, M_4 , the dynamic effect is *sustained* and lasts for the complete test length. All previous motive related responses are added up in the sustained dynamic effect. This model represents the idea that motive strength does not recover during the test administration.

Statistical details. In the following part of the paper we describe the statistical details and formulations of the Thurstonian IRT models discussed earlier. We first describe the level 1 specifications that reflects the classic Thurstonian model that does not differentiate between people. Then we describe the level 2 specifications that add person parameters (M_1), picture parameters (M_2), and the dynamic effects (M_3 and M_4). The model formulation for the level 2 specifications are provided in Table 1.

The probability that person s prefers one motive category (*aff*, *ach*, *pow*, and *zero* for other content) to another category can be written as $\pi_{affachs}$, $\pi_{affpows}$, $\pi_{affzeros}$, $\pi_{achpows}$, $\pi_{achzeros}$, and $\pi_{powzeros}$. These are a function of differences between the mean evaluations (latent utilities) of motive categories. The mean evaluations are denoted as μ_{affs} and μ_{achs} , μ_{pows} , μ_{zeros} , and are linked to the binomial outcome using a probit link. The probit model can then be written as:

$$\begin{pmatrix} \text{probit}(\pi_{affachs}) \\ \text{probit}(\pi_{affpows}) \\ \text{probit}(\pi_{affzeros}) \\ \text{probit}(\pi_{achpows}) \\ \text{probit}(\pi_{achzeros}) \\ \text{probit}(\pi_{powzeros}) \end{pmatrix} = \begin{pmatrix} 1 & -1 & 0 \\ 1 & 0 & -1 \\ 1 & 0 & 0 \\ 0 & 1 & -1 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \mu_{affs} \\ \mu_{achs} \\ \mu_{pows} \\ 0 \end{pmatrix} = \mathbf{D}\mu_s.$$

The design matrix \mathbf{D} consists of four dummy variables that refer to the four categories (the three motives and zero motive responses) and is multiplicatively related to the vector of mean evaluations. The last row of dummy variables can be omitted without a loss of generality and the equivalent row of mean evaluations is constrained to zero. This probit model does not include an intercept.

Model M_1 adds random effects ν to the motive categories to capture the variance of person s for the respective motive, they are normally distributed around 0. The model formulation is provided in Table 1.

Model M_2 adds a picture covariate x_{ms} ($m = 1, \dots, M$) with M being the total number of pictures. If person s writes a story with motive content for the affiliation motive in picture m then $x_{affms} = 1$, if the response shows no motive content for the affiliation motive, then $x_{affms} = 0$. The mean evaluation of picture m for the corresponding motive category is denoted as β_m .

In applying Model M_2 to the OMT, it is important to consider some specific characteristics of this test. The OMT includes a larger number of pictures. However, some pictures in the OMT evoke stories with content from mostly one or two motives. For the other motives, the response probabilities can then be very low. IRT models that are effectively a version of the generalized linear mixed-effects model generally have difficulties to estimate effects when response probabilities are very low because the information included in these responses is limited which leads to complete separations in model estimation (Fox & Weisberg, 2011; Heinze & Schemper, 2002). The standard approach in the IRT literature is typically to exclude stimuli with extremely low or high response probabilities unless the researcher has a strong theoretical reason to assume that the stimuli actually measures the characteristic of interest (Debeer & Janssen, 2013; Wainer, Bradlow, & Wang, 2007; Wright & Stone, 1979). For the OMT, it is theoretically reasonable to assume that some pictures simply do not invoke certain motives. Therefore, it typically makes sense to not include pairwise comparisons for specific motives and pictures that show low response probabilities when estimating the respective latent motive.

A temporary dynamic reduction in motive strength is modelled in M_3 . Therefore, another fixed effect TDE (temporary dynamic effect) can be added to model M_2 for each motive. TDE_s is simply denoting the response of the previous picture for the respective motive (0 = no response in the previous picture, 1 = one response in the previous picture).

M_4 models the sustained dynamic effect. A fixed effect SDE (sustained dynamic effect) is used instead of the temporary dynamic effect TDE of model M_3 . SDE_s is simply denoting the total number of previous responses for the respective motive (0 = no previous response, 1 = one previous response, 2 = two previous responses, ...).

Estimation Methods. Thurstonian IRT models can be estimated using broadly available software designed to model multilevel data with binary outcomes (generalized linear mixed-effects models; Böckenholt, 2004, 2006). In this paper we demonstrate the implementation of these models in the R environment (R Core Team, 2015) using the `glmer` function for general linear mixed-effect models (Bates, Mächler, Bolker, & Walker, 2015) in the `lme4` 1.1-12 package. The `glmer` function uses the Laplace approximation. However, other frequentist or Bayesian software packages for generalized linear-mixed effects models that allow for multiple random effects are equally capable to estimate these types of models. For instance, Lang (2014) also used Bayesian generalized linear mixed-effects models as implemented in the software package `MCMCglmm` (Hadfield, 2010). Furthermore, Thurstonian models can also be implemented using structural equation modeling software (A. Brown & Maydeu-Olivares, 2011). From our experience, generalized linear mixed-effects models software can more easily deal with missing pairwise comparisons for individual responses because it uses a long format. As one response to an OMT picture has always three missing pairwise comparisons (“take the best”

response format), generalized linear mixed-effects models software may therefore be more suitable for the OMT.

IRT Reliability

One effective strategy to understand reliability in complex IRT models is to use simulation studies (A. Brown & Maydeu-Olivares, 2011; Lang, 2014; Reise & Yu, 1990; Weiss, 1982). The goal of this approach is to investigate how well model parameters and underlying latent dispositions can be recovered from simulated data by simulating the entire response process. Such a simulation study consists of three steps. In the first step, one simulates datasets generated on basis of the empiric sample and the Thurstonian IRT model. The basis for the simulated datasets are the estimated model parameters from the empiric dataset which are used as the underlying true population estimates. In the second step, the IRT model is fitted to these simulated datasets. In the third step, one correlates the true population estimates with the estimates that one gets from fitting the simulated data. The goal of this step is to examine how well the IRT model is able to recover the true population estimates from the simulated data. Correlations between true motive scores and simulated IRT scores are called fidelity correlations (Weiss, 1982). Squared fidelity correlations equal the squared correlation reliability, which is known as a common definition of the reliability (Lord, Novick, & Birnbaum, 1968). We first run the simulation 100 times to get average fidelity correlations. It is then necessary to repeat the procedure again and correct for potential systematic estimation bias, which is commonly necessary for complex and multidimensional IRT models (Kuk, 1995; Ng, Carpenter, Goldstein, & Rasbash, 2006). For the bias correction, we first identify the bias as the difference between the actual picture parameters from the empirical sample and the average picture parameters from the first 100 simulation runs. In the bias corrected 100 simulation runs we subtract this bias from the

empiric picture parameters used in the first step of this procedure and thereby correct for the bias.

Empirical Study

To evaluate the viability of applying the dynamic Thurstonian IRT approach (Lang, 2014) to the OMT, we analyze a dataset composed of three smaller samples (sample one $n=250$; sample two $n = 275$; sample three $n = 112$). We combined data to one larger set, because psychometric analysis become more robust with a higher number of respondents. Subsets of two of the samples have been published earlier (sample one: Schöler et al. (2015); sample two: Lang et al. (2012)).

The empirical study has two goals. We first want to investigate if the models fit our data and evaluate which model fits best. Based on previous research (Lang, 2014) we expect model M_4 , which accounts for motive competition and sustained dynamic effects, to show the best model fit. Second, we want to study the IRT reliability of the OMT and compare it to traditional reliability estimates. As the Thurstonian IRT model accounts for underlying response processes, we expect to find higher reliability coefficients than in an analysis of the sum scores with Cronbach's alpha or the greatest lower bound, which is less conservative than Cronbach's alpha in estimating reliability (Sijtsma, 2008).

Method

Participants. From the overall sample of 637, four participants with more than five not motive related or missing codings were excluded from the analysis. This approach is typically recommended to exclude participants who did not understand the test instructions and is also in line with the measure recommendations in the manual. The final sample of 633 OMTs comprises

403 women and 220 men (10 participants did not indicate their gender) with an age range from 18 to 69 ($M = 31,71$, $SD = 12,43$; 16 participants did not indicate their age) years.

Measures. Six coders evaluated the sample of 637 OMTs, using detailed information provided by the OMT manual (Kuhl & Scheffer, 1999, 2002). In addition, the coders were instructed and trained through various OMT coding seminars. In order to evaluate interrater reliability between the coders we calculated Gwet's AC_1 statistic (Gwet, 2008a, 2008b), which is an improved version of Cohen's κ (kappa). Coder A and B coded 250 OMTs from the first sample and showed good interrater reliability (0.98 for affiliation, 0.97 for achievement, and .87 for power). A third coder C has coded the remaining 275 OMTs from the second sample. To ensure coding quality, 165 OMTs coded by coder C have also been coded by a coder D, showing again sufficient interrater reliability (0.88 for affiliation, 0.87 for achievement, and 0.76 for power). Coder E and F coded 112 OMTs from the third sample and showed sufficient interrater reliability (0.92 for affiliation, 0.90 for achievement, and .83 for power). Disagreements between coders were solved through discussions in additional sessions in the first sample. The codings of the main coders C and E were used in samples two and three. To get the most distinct approach category, we integrated the first three OMT categories for each motive into one approach factor.

Results

Table 2 shows descriptive information for the OMT picture set. Pictures clearly differ in their activation strength for the motives. Some pictures show very few responses for a particular motive. As discussed earlier, we excluded motive comparisons that had a low response probability (lower than 1%). We eliminated the comparisons for specific motives in seven pictures (pictures 1,2,4,6,7,8 and 15; see Table 5) in models M_2 , M_3 and M_4 .

Table 3 shows the fit statistics for all model tests. A lower log-likelihood indicates a better model fit. To analyze the first question (which model shows the best fit) we started by comparing the model fit of M_1 , the most basic model, to the fit of M_2 , that additionally accounts for picture effects. Picture effects seem to improve model fit. M_2 shows better model fit in comparison with M_1 . This is in line with the observation of the descriptive information given in Table 2 (i.e., motive scores differ substantially between pictures). The model fit improved further by adding dynamic effects (M_3 and M_4). Both models with dynamic effects improved model fit, but the extent to which they improved model fit differs. A direct test of difference in fit for the two dynamic Thurstonian models is not possible because the models are not nested. However, the comparison of all information indices for the two models with different dynamic effects (M_3 and M_4) shows that model M_4 (sustained dynamic effects) fits best. All three sustained dynamic effects are negative (Table 5), which confirms our expectation of a dynamic implicit motive system and is in line with existing literature (Lang, 2014). Because M_4 fits best to the data, we used it as the standard model to run all following analysis.

Table 4 provides the IRT reliability estimates from the simulation study. Affiliation, achievement, and power were $r_{\theta\theta}^2 = .52$, $r_{\theta\theta}^2 = .62$, and $r_{\theta\theta}^2 = .73$ respectively. Table 4 also provides Cronbach's alpha and greatest lower bound. Both were considerably lower, in line with our expectations.

Table 5 shows picture effects for model M_4 from the empiric sample and from the simulated datasets as well as the corrected simulated dataset. They indicate how likely it is to write a story with respective motive contents. For example, in picture 9, an achievement related response is more likely than an affiliation related response and an affiliation related response is more likely than a power related response. Table 5 further provides the sustained dynamic effects

SDE. A negative value indicates that a previous story with this motive content lowers the probability of another story with the same motive content.

Table 5 also provides correlations between the IRT scores for each motive, which were low. Correlations between the sum and IRT θ scores were $r(631) = .96$ for affiliation, $r(631) = .96$ for achievement and $r(631) = .96$ for power.

Tutorial/Demonstration: Fitting Dynamic Thurstonian IRT Models Using Generalized Linear Mixed Effects Modeling Software

We provide a generated OMT dataset to fit the models in order enable researchers to reproduce the application of our dynamic Thurstonian IRT model on OMT data as online supplemental material (“OMT generated dataset”). It is also possible to use the script in Appendix A, that generates the OMT dataset as the first step of the reliability estimation. Note that the OMT generated dataset as well as the script in Appendix A are based on the parameters of the model with the best fit, M_4 . It is, however, possible to fit the other models to these data for demonstrational purposes.

Fitting the Thurstonian Models

Recoding the data. As described above, one receives information about three pairwise comparisons for each motive response in the OMT. Therefore, it is necessary to recode the OMT data first. One needs three rows in the OMT data matrix for a response of one person to a particular picture. Three dummy variables (DAFF, DACH, DPOW) indicate the three motives. Table 6 shows an example of an OMT data matrix recoded into a pairwise comparison matrix. The first picture is coded as affiliation. The first row represents the comparison between affiliation and achievement, the second row between affiliation and power, and the third row between affiliation and the zero category. Similarly, picture two is an example for an

achievement response, picture three is an example for a power response and picture four is an example for a zero response. Each picture consists of three rows representing the three comparisons. The dynamic effects are also specified in the recoded OMT data matrix. TDACH, TDAFF and TDPOW indicate the temporary dynamic effect. In the sample matrix presented in Table 6, the first picture is an affiliation response. It follows that TDAFF is 1 for the second picture. The second picture is an achievement response, so that TDACH is 1 for the third picture – TDAFF resets to 0. SDAFF, SDACH, and SDPOW indicate all potential previous responses of the respective motive, they reflect the sustained dynamic effects. In the sample matrix, the first picture is an affiliation response. It follows that SDAFF is 1 for the second picture. The second picture is an achievement response, so that SDACH is 1 for the third picture, SDAFF is still 1.

Fitting the data. After recoding the data, models M_1 , M_2 , M_3 and M_4 can be fitted using the lme4 1.1-12 package in the R environment (R Core Team, 2015). The lme 4 package fits linear mixed-effect models with the glmer function. The glmer function is a two-sided linear formula object with the response variables `response1` and `response2` on the left side of the `~` operator and the terms of the model on the right side, separated by `+` operators. For model M_1 , latent motive traits are modelled as random effects and written in parenthesis. Vertical bars separate the expressions for the design matrices from the grouping factor `persons`:

```
glmer(cbind(response1, response2) ~
0
+ (0+DAFF+DACH+DPOW|person),
family=binomial("probit"), tdata,
control=glmerControl(optimizer="nloptwrap", optCtrl=list(maxfun=1e10), calc.derivs=F))2
```

In model M_2 , fixed effects are added for the pictures:

```
tdata$PICAFF<-as.factor(ifelse(tdata$DAFF!=0,tdata$picture,1))
tdata$PICACH<-as.factor(ifelse(tdata$DACH!=0,tdata$picture,3))
tdata$PICPOW<-as.factor(ifelse(tdata$DPOW!=0,tdata$picture,1))

glmer(cbind(response1, response2) ~
0
```

```
+ (0+DAFF+DACH+DPOW|person)
+PICAFF:DAFF
+PICACH:DACH
+PICPOW:DPOW,
family=binomial("probit"), tdata,
control=glmerControl(optimizer="nloptwrap", optCtrl=list(maxfun=1e10), calc.derivs=F))2
```

In model M_3 , temporary dynamic effects are added:

```
tdata$PICAFF<-as.factor(ifelse(tdata$DAFF!=0,tdata$picture,1))
tdata$PICACH<-as.factor(ifelse(tdata$DACH!=0,tdata$picture,3))
tdata$PICPOW<-as.factor(ifelse(tdata$DPOW!=0,tdata$picture,1))

M3<-glmer(cbind(response1, response2) ~
0
+ (0+DAFF+DACH+DPOW|person)
+PICAFF:DAFF
+PICACH:DACH
+PICPOW:DPOW
+DAFF:TDAFF+DACH:TDACH+DPOW:TDPOW,
family=binomial("probit"), tdata,
control=glmerControl(optimizer="nloptwrap", optCtrl=list(maxfun=1e10), calc.derivs=F))2
```

In model M_4 , sustained dynamic effects are added instead of the temporary dynamic effects from

M_3 :

```
tdata$PICAFF<-as.factor(ifelse(tdata$DAFF!=0,tdata$picture,1))
tdata$PICACH<-as.factor(ifelse(tdata$DACH!=0,tdata$picture,3))
tdata$PICPOW<-as.factor(ifelse(tdata$DPOW!=0,tdata$picture,1))

glmer(cbind(response1, response2) ~
0
+ (0+DAFF+DACH+DPOW|person)
+PICAFF:DAFF
+PICACH:DACH
+PICPOW:DPOW
+DAFF:SDAFF+DACH:SDACH+DPOW:SDPOW,
family=binomial("probit"), tdata,
control=glmerControl(optimizer="nloptwrap", optCtrl=list(maxfun=1e10), calc.derivs=F))2
```

The `glmer` function uses the Laplace approximation (Böckenholt, 2001) and estimates random effects with the maximum a posteriori method (De Boeck et al., 2011). In the empiric demonstration of the model we use the `bobyqa` optimizer from the `nloptr` package in R, because this optimizer is typically the fastest. The substantive conclusions did not differ using other `lme4` optimizers (Bates et al., 2015).

Estimating IRT Reliability

In this part of the paper we describe the implementation of the simulation approach to estimate the IRT reliability of the OMT in the R environment. As described above, the approach determines the degree to which the model can recover true scores from simulated datasets. The approach includes three steps.

First Step. The goal of the first step is to simulate a dataset based on the empirical model. The estimated model parameters from the empirical sample are used as the underlying true population estimates to generate the new dataset. The code for this step is provided in Appendix A. The estimated model parameters used in this demonstration are from model M_4 and the empirical study that we describe in the applied part.

In the `tv` object we specify the following parameters from our empirical OMT sample. The `pictureAff`, `pictureAch`, and `picturePow` vectors contain the information about the empiric picture parameters. `NA` indicates that this picture does not measure the respective motive and reflects the resolution of low response probabilities. The `de` vector contains the fixed dynamic effects, the `ranefsd` vector stores the standard deviations of the random effects, which are the estimated latent traits, and the `ranefcor` object contains the correlations between those latent traits. `n` describes the number of respondents.

The `genomtdat` function generates the dataset based on the parameters stored in the `tv` object. We use the `mvtnorm` package in R (Genz et al., 2016), that was made to generate multivariate normal and t distributions. The resulting simulated dataset is then stored in the `tdata` matrix. The matrix is similar to the matrix provided in Table 6.

Second Step. In the second step, the IRT model, in this demonstration M_4 , is fitted to the simulated dataset. The code is identical to the code provided above for the fit of model M_4 .

Third Step. In the third step, one examines how well the IRT model is able to recover the true underlying latent motives with model M_4 from the simulated datasets. Therefore, we correlate the estimated latent trait scores from the simulated dataset—the outcome of the second step—with the latent trait scores from the empiric sample. The square of this correlation is called fidelity correlation (Weiss, 1982) and is equal to the squared correlation reliability, a common definition of reliability (Lord et al., 1968).

```
diag(cor(ranef(mod1)$person, unique(tdata[,2:4]))^2)
```

`ranef(mod1)$person` indicates the estimated latent trait scores from the simulated dataset and `unique(tdata[,2:4])` refers to the latent trait scores from the empiric sample.

One then simulates 100 datasets and repeats the procedure twice to correct for potential systematic estimation biases as described in the introduction.

Discussion

Researchers have long discussed the measurement properties of implicit motive measures (e.g., Entwisle, 1972) and in recent years have made significant progress in understanding the response processes in these types of measures (Tuerlinckx et al., 2003; Lang et al., 2012; Lang, 2014). The present paper contributes to this emerging line of research. The specific goals of our study were to (a) apply recently developed dynamic Thurstonian IRT models (Lang, 2014) to the OMT and to (b) demonstrate the use of these models with OMT data. We found that the response process in the OMT can be described using a Thurstonian IRT model with a sustained dynamic effect and this study thus extends and replicates earlier findings. Like in previous research, we also found that IRT reliabilities were much higher than internal consistency estimates. The IRT

reliability for power reached .70 and the reliability for achievement was still in a range that is typically sufficient for research purposes (Ellis, 2013).

Theoretical Implications

This research has two theoretical implications for the understanding of implicit motive measures. One important finding of our study is that the Thurstonian IRT model with a sustained dynamic effect describes the response process in implicit motives measures better than the alternative Thurstonian models we examined. This finding thus generalizes from the original study (Lang, 2014). A possible implication of this finding is that dynamic processes in implicit motive measures capture a fundamental mechanism in the human motivation system. Dynamic processes have always been an important assumption in dynamic system models of motivation (Atkinson & Birch, 1970; Atkinson et al., 1977; Revelle, 1986). Dynamic Thurstonian IRT makes it possible to study and model these processes. We suggest that these models could possibly also serve as a general tool to study motivational behavior. For instance, the model could be applied to social interactions in laboratory situations or be used to study motive change in diary studies to understand daily motivational processes.

This study also contributes to a deeper understanding of the exact nature of dynamic motivational processes in implicit motive measures. Atkinson and Birch (1970) suggested that motives are temporally reduced by action for a limited period, so that motivational behavior typically can be described by a saw tooth pattern (also see Carver & Scheier, 2002; Wright, 2016). This study replicates previous research (Lang, 2014) suggesting that the dynamic Thurstonian model in which consumption of motive strength is sustained (M_4) provides a better fit than a saw tooth model (M_3). Implications of this finding is that motives have a greater inertia than originally assumed by theoretical work (Atkinson, 1950; Atkinson et al., 1977; Reitman &

Atkinson, 1958). A reviewer noted that the two models we examined are not the only possible dynamic models and that in-between models are possible. Specifically, it is possible that the dynamic effects last longer than one picture (M_3) but shorter than the duration of the test (M_4). We therefore also fitted two other models, in which the dynamic effects lasted not only one but two or three pictures, before motive strength recovers. The model with a dynamic effect that lasted two pictures (deviance = 19,957; AIC = 20,047; BIC = 20,408) and the model with a dynamic effect that lasted three pictures (deviance = 19,924, AIC = 20,014, BIC = 20,375), had a somewhat better fit than M_3 (deviance = 19,970; AIC = 20,060; BIC = 20,422) but did not fit as well as M_4 (deviance = 19,204; AIC = 19,294, BIC = 19,656). Another possibility is that a general fatigue model in which the dynamic effect is not motive-specific could possibly fit the data equally good or better than M_4 . We tested a model, in which the response probability for all three motives declines with each motive-related response to a picture no matter what type of motive-related response occurred. This model did not provide a better fit than M_4 (deviance = 19,964; AIC = 20,054; BIC = 20,415).

Measurement Implications

Our research has three measurement implications for research on implicit motives. One important implication of our research is that the OMT is an implicit motive measure that can be used to measure implicit motives in research. Although the IRT reliability estimates for affiliation and achievement in our study were smaller than the minimum requirements for questionnaire measures typically recommended in the literature (e.g., Nunnally, 1978), estimates around .60 are typically sufficient for research purposes (Ellis, 2013). The OMT is a procedure that is relatively easy to learn because coders do not need to separate texts in different parts and only one motive response for each picture is coded. The OMT may therefore have practical

advantages for research purposes. The Thurstonian approach suggests that it is beneficial to present pictures that stimulate multiple motives and code those motives to assess the motive of interest even if one is only interested in a single motive. This idea is also in line with other recent recommendations in the literature (Ramsay & Pang, 2013).

A second measurement implication of our study is that researchers should not interpret internal consistency estimates for the OMT. As our analyses show, Cronbach's alpha, as well as the greatest lower bound (Woodhouse & Jackson, 1977), likely do not adequately estimate the true underlying reliability of the response process.

A third measurement implication of our study is that sum score estimates based on OMT responses are likely useful for research and can be used in future research. We correlated sum scores and IRT θ scores from the dynamic Thurstonian IRT model and our results suggested that these estimates are very high. Researchers who seek to use the OMT and do not have sample sizes available that allow them to estimate IRT models, can likely confidentially use the OMT sum scores as an approximation of IRT θ .

Limitations and Future Directions

One limitation of the OMT that has become apparent in our study is the fact that some pictures do not capture some of the motives, which shortens test length. From a psychometric perspective, one potential way to further develop the OMT and increase its IRT reliability may be to include more pictures that simultaneously capture all three motives and thereby increase test length. Including ambiguous pictures in measures of implicit motives has previously been recommended (Murstein, 1965; Pang, 2010; Smith, 1992). However, it is possible that the pictures that only capture a specific motive do so particularly well. It is likely useful to examine the implications of different types of pictures for motive measurement.

Another limitation of the OMT and other implicit motive measures is the fact that these measures cannot be lengthened to an unlimited amount (Reitman & Atkinson, 1958; Schultheiss & Pang, 2007). The dynamic Thurstonian model provides a formal theory on why this is the case and in so doing also demonstrates a natural limitation of measurement procedures like the OMT. We nevertheless believe that future research could systematically investigate to what degree the OMT can be lengthened and at what point the dynamic effects become so strong that adding additional pictures is not useful anymore.

References

- Atkinson, J. W. (1950). Studies in projective measurement of achievement motivation. In *Microfilm Abstracts*.
- Atkinson, J. W., & Birch, D. (1970). *Dynamics of Action*. New York: Wiley.
- Atkinson, J. W., Bongort, K., & Price, L. H. (1977). Explorations using computer simulation to comprehend thematic apperceptive measurement of motivation. *Motivation and Emotion*, 1, 1–27. <https://doi.org/10.1007/BF00997578>
- Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 1. <https://doi.org/10.18637/jss.v067.i01>
- Baumann, N., Kaschel, R., & Kuhl, J. (2005). Striving for Unwanted Goals: Stress-Dependent Discrepancies Between Explicit and Implicit Achievement Motives Reduce Subjective Well-Being and Increase Psychosomatic Symptoms. *Journal of Personality and Social Psychology*, 89, 781–799. <https://doi.org/10.1037/0022-3514.89.5.781>
- Baumann, N., Kaschel, R., & Kuhl, J. (2010). Implicit Motives: A Look from Personality Systems Interaction Theory. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit Motives*. New York: Oxford University Press.
- Baumann, N., & Scheffer, D. (2010). Seeing and mastering difficulty: The role of affective change in achievement flow. *Cognition & Emotion*, 24, 1304–1328. <https://doi.org/10.1080/02699930903319911>
- Böckenholt, U. (2001). Hierarchical Modeling of Paired Comparison Data. *Psychological Methods*, 6, 49–66. <https://doi.org/10.1037//1082-989X.6.1.49>
- Böckenholt, U. (2004). Comparative judgments as an alternative to ratings: identifying the scale origin. *Psychological Methods*, 9, 453–65. <https://doi.org/10.1037/1082-989X.9.4.453>

- Böckenholt, U. (2006). Thurstonian-Based Analyses: Past, Present, and Future Utilities. *Psychometrika*, 71, 615–629. <https://doi.org/10.1007/s11336-006-1598-5>
- Boeck, P. de. (2008). Random Item IRT Models. *Psychometrika*, 73, 533–559. <https://doi.org/10.1007/s11336-008-9092-x>
- Brown, A., & Maydeu-Olivares, A. (2011). Item Response Modeling of Forced-Choice Questionnaires. *Educational and Psychological Measurement*, 71, 460–502. <https://doi.org/10.1177/0013164410375112>
- Brown, A., & Maydeu-Olivares, A. (2013). How IRT can solve problems of ipsative data in forced-choice questionnaires. *Psychological Methods*, 18, 36–52. <https://doi.org/10.1037/a0030641>
- Carver, C. S., & Scheier, M. F. (2002). Control processes and self-organization as complementary principles underlying behavior. *Personality and Social Psychology Review*, 6(4), 304–315.
- Critchlow, D. E., & Fligner, M. A. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation on GLIM. *Psychometrika*, 56, 517–533. <https://doi.org/10.1007/BF02294488>
- Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika*, 16, 297–334. <https://doi.org/10.1007/BF02310555>
- De Boeck, P., Bakker, M., Zwitser, R., Nivard, M., Hofman, A., Tuerlinckx, F., & Partchev, I. (2011). The Estimation of Item Response Models with the lmer Function from the lme4 Package in R. *Journal Of Statistical Software*, 39, 1–28.
- Debeer, D., & Janssen, R. (2013). Modeling Item-Position Effects Within an IRT Framework. *Journal of Educational Measurement*, 50, 164–185.

- Doran, H., Bates, D., Bliese, P., & Dowling, M. (2007). Estimating the Multilevel Rasch Model : With the lme4 package. *Journal of Statistical Software*, 20, 1–18.
<https://doi.org/10.1111/j.1467-9868.2007.00600.x>
- Ellis, J. L. (2013). A standard for test reliability in group research. *Behavior Research Methods*, 45, 16–24. <https://doi.org/10.3758/s13428-012-0223-z>
- Emmons, R. A. (1989). Exploring the Relations Between Motives and Traits: The Case of Narcissism. In D. Buss & N. Cantor (Eds.), *Personality Psychology SE - 3* (pp. 32–44). Springer US. https://doi.org/10.1007/978-1-4684-0634-4_3
- Entwisle, D. R. (1972). To Dispel Fantasies About Fantasy-Based Measures of Achievement Motivation. *Psychological Bulletin*, 77, 377–391. <https://doi.org/10.1037/h0020021>
- Fineman, S. (1977). The achievement motive construct and its measurement: Where are we now? *British Journal of Psychology*, 68, 1–22. <https://doi.org/10.1111/j.2044-8295.1977.tb01554.x>
- Fox, J., & Weisberg, S. (2011). *An R Companion to Applied Regression*. SAGE Publications.
Retrieved from <https://books.google.be/books?id=YH6NotdvzF0C>
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., & Hothorn, T. (2016). mvtnorm: Multivariate Normal and t Distributions. Retrieved from <http://cran.r-project.org/package=mvtnorm>
- Gwet, K. L. (2008a). Computing inter-rater reliability and its variance in the presence of high agreement. *The British Journal of Mathematical and Statistical Psychology*, 61, 29–48.
<https://doi.org/10.1348/000711006X126600>
- Gwet, K. L. (2008b). Variance Estimation of Nominal-Scale Inter-Rater Reliability with Random Selection of Raters. *Psychometrika*, 73, 407–430. <https://doi.org/10.1007/s11336-007-9054->

8

- Hadfield, J. D. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, 1(2).
<https://doi.org/10.18637/jss.v033.i02>
- Hall, J. L., Stanton, S. J., & Schultheiss, O. C. (2010). Biopsychosocial and Neural Processes of Implicit Motivation. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit Motives* (pp. 279–307). New York: Oxford University Press.
- Heinze, G., & Schemper, M. (2002). A solution to the problem of separation in logistic regression. *Statistics in Medicine*, 21(16), 2409–2419. <https://doi.org/10.1002/sim.1047>
- Hofer, J., Busch, H., Chasiotis, A., Kärtner, J., & Campos, D. (2008). Concern for Generativity and Its Relation to Implicit Pro-Social Power Motivation, Generative Goals, and Satisfaction With Life: A Cross-Cultural Investigation. *Journal of Personality*, 76, 1–30.
<https://doi.org/10.1111/j.1467-6494.2007.00478.x>
- Kazén, M., & Kuhl, J. (2005). Intention memory and achievement motivation: volitional facilitation and inhibition as a function of affective contents of need-related stimuli. *Journal of Personality and Social Psychology*, 89, 426–448. <https://doi.org/10.1037/0022-3514.89.3.426>
- Kazén, M., & Kuhl, J. (2011). Directional discrepancy between implicit and explicit power motives is related to well-being among managers. *Motivation and Emotion*, 35, 317–327.
<https://doi.org/10.1007/s11031-011-9219-8>
- Köllner, M. G., & Schultheiss, O. C. (2014). Meta-analytic evidence of low convergence between implicit and explicit measures of the needs for achievement, affiliation, and power. *Frontiers in Psychology*, 5, 826. <https://doi.org/10.3389/fpsyg.2014.00826>

- Kuhl, J., & Kazén, M. (2008). Motivation, affect, and hemispheric asymmetry: power versus affiliation. *Journal of Personality and Social Psychology*, 95, 456–469.
<https://doi.org/10.1037/0022-3514.95.2.456>
- Kuhl, J., & Scheffer, D. (1999). Auswertungsmanual für den Operanten Motiv-Test (OMT).
- Kuhl, J., & Scheffer, D. (2002). Der operante Multi-Motiv-Test (OMT): Manual [The operant multi-motive-test (OMT): Manual]. Osnabrück, Germany: University of Osnabrück.
- Kuk, Y. C. A. (1995). Asymptotically unbiased estimation in generalized linear models with random effects. *Journal of the Royal Statistical Society: Series B (Methodological)*, 57, 395–407.
- Lang, J. W. B. (2014). A dynamic Thurstonian item response theory of motive expression in the picture story exercise: Solving the internal consistency paradox of the PSE. *Psychological Review*, 121, 481–500. <https://doi.org/10.1037/a0037011>
- Lang, J. W. B., Zettler, I., Ewen, C., & Hülshager, U. R. (2012). Implicit motives, explicit traits, and task and contextual performance at work. *Journal of Applied Psychology*, 97, 1201–1217. <https://doi.org/10.1037/a0029556>
- Lilienfeld, S. O., Wood, J. M., & Garb, H. N. (2000). The Scientific Status of Projective Techniques. *Psychological Science in the Public Interest*, 1, 27–66.
<https://doi.org/10.1111/1529-1006.002>
- Lord, F. M., Novick, M. R., & Birnbaum, A. (1968). *Statistical theories of mental test scores*. Reading, MA: Addison-Wesley.
- McAdams, D. P., Jackson, R. J., & Kirshnit, C. (1984). Looking, laughing, and smiling in dyads as a function of intimacy motivation and reciprocity. *Journal of Personality*, 52, 261–273.
<https://doi.org/10.1111/j.1467-6494.1984.tb00881.x>

- McClelland, D. C. (1987). *Human motivation*. CUP Archive.
- McClelland, D. C., Atkinson, J. W., Clark, R. A., & Lowell, E. L. (1953). The measuring instrument. In *The achievement motive* (pp. 185–217). CHAP, East Norwalk, CT, US: Appleton-Century-Crofts. <https://doi.org/10.1037/11144-007>
- McClelland, D. C., & Boyatzis, R. E. (1982). Leadership motive pattern and long-term success in management. *Journal of Applied Psychology*, 67, 737–743. <https://doi.org/10.1037/0021-9010.67.6.737>
- McClelland, D. C., Koestner, R., & Weinberger, J. (1989). How do self-attributed and implicit motives differ? *Psychological Review*, 96, 690–702. <https://doi.org/10.1037/0033-295X.96.4.690>
- Morgan, C. D., & Murray, H. A. (1935). A method for investigating fantasies: The Thematic Apperception Test. *Archives of Neurology & Psychiatry*, 34, 289–306. <https://doi.org/10.1001/archneurpsyc.1935.02250200049005>
- Murray, H. A. (1938). *Explorations in Personality; A Clinical and Experimental Study of Fifty Men of College Age*. *American Sociological Review*. New York: Oxford University Press. <https://doi.org/10.2307/2084329>
- Murstein, B. I. (1965). Projection of hostility on the TAT as a function of stimulus, background, and personality variables. *Journal of Consulting Psychology*, 29, 43–48. <https://doi.org/10.1037/h0021666>
- Ng, E. S. W., Carpenter, J. R., Goldstein, H., & Rasbash, J. (2006). Estimation in generalised linear mixed models with binary outcomes by simulated maximum likelihood. *Statistical Modelling*, 6, 23–42. <https://doi.org/10.1191/1471082X06st106oa>
- Nunnally, J. C. (1978). *Psychometric Theory*. McGraw-Hill Book Company.

- Pang, J. S. (2010). Content coding methods in implicit motive assessment: Standards of measurement and best practices for the picture story exercise. In O. C. Schultheiss & J. C. Brunstein (Eds.), *Implicit Motives* (pp. 119–1550). New York: Oxford University Press.
- R Core Team. (2015). R: A Language and Environment for Statistical Computing. *R Foundation for Statistical Computing*, v3.0.1, 3604. <https://doi.org/10.1007/978-3-540-74686-7>
- Ramsay, J. E., & Pang, J. S. (2013). Set ambiguity: A key determinant of reliability and validity in the picture story exercise. *Motivation and Emotion*, 37, 661–674. <https://doi.org/10.1007/s11031-012-9339-9>
- Reise, S. P., & Yu, J. (1990). Parameter Recovery in the Graded Response Model Using MULTILOG. *Journal of Educational Measurement*, 27, 133–144. <https://doi.org/10.1111/j.1745-3984.1990.tb00738.x>
- Reitman, W. R., & Atkinson, J. W. (1958). Some methodological problems in the use of thematic apperceptive measures of human motives. In J. W. Atkinson (Ed.), *Motives in fantasy, action, and society* (pp. 664–683). Princeton, NJ: Van Nostrand.
- Revelle, W. (1986). Motivation and Efficiency of Cognitive Performance BT - Frontiers of Motivational Psychology: Essays in Honor of John W. Atkinson. In D. R. Brown & J. Veroff (Eds.) (pp. 107–131). New York, NY: Springer US. https://doi.org/10.1007/978-1-4684-6341-5_7
- Revelle, W. (2016). psych: Procedures for Psychological, Psychometric, and Personality Research. Evanston, Illinois. Retrieved from <https://cran.r-project.org/package=psych>
- Scheffer, D. (2005). *Implizite Motive [Implicit motives]*. Göttingen, Germany: Hogrefe.
- Schüler, J., Brandstätter, V., Wegner, M., & Baumann, N. (2015). Testing the convergent and discriminant validity of three implicit motive measures: PSE, OMT, and MMG. *Motivation*

and Emotion, 39, 839–857. <https://doi.org/10.1007/s11031-015-9502-1>

Schüler, J., Job, V., Fröhlich, S. M., & Brandstätter, V. (2008). A high implicit affiliation motive does not always make you happy: A corresponding explicit motive and corresponding behavior are further needed. *Motivation and Emotion*, 32, 231–242.

<https://doi.org/10.1007/s11031-008-9096-y>

Schüler, J., Job, V., Fröhlich, S. M., & Brandstätter, V. (2009). Dealing with a 'hidden stressor': Emotional disclosure as a coping strategy to overcome the negative effects of motive incongruence on health. *Stress and Health*, 25, 221–233. <https://doi.org/10.1002/smi.1241>

Schultheiss, O. C. (2008). Implicit motives. In O. P. John, R. W. Robins, & L. A. Pervin (Eds.), *Handbook of personality: Theory and research* (3rd ed., pp. 603–633). New York: The Guilford Press.

Schultheiss, O. C., Lienen, S. H., & Schad, D. (2008). The reliability of a Picture Story Exercise measure of implicit motives: Estimates of internal consistency, retest reliability, and ipsative stability. *Journal of Research in Personality*, 42, 1560–1571.

<https://doi.org/10.1016/j.jrp.2008.07.008>

Schultheiss, O. C., & Pang, J. S. (2007). Measuring implicit motives. In R. W. Robins, R. C. Fraley, & R. Krueger (Eds.), *Handbook of Research Methods in Personality Psychology* (pp. 322–344). New York: Guilford.

Sijtsma, K. (2008). On the Use, the Misuse, and the Very Limited Usefulness of Cronbach's Alpha. *Psychometrika*, 74, 107–120. <https://doi.org/10.1007/s11336-008-9101-0>

Smith, C. P. (1992). *Motivation and personality: Handbook of thematic content analysis*. New York: Cambridge University Press.

Spangler, W. D. (1992). Validity of questionnaire and TAT measures of need for achievement:

- Two meta-analyses. *Psychological Bulletin*, 112, 140–154. <https://doi.org/10.1037/0033-2909.112.1.140>
- Thurstone, L. L. (1927). A law of comparative judgment. *Psychological Review*, 34, 273–286. <https://doi.org/10.1037/h0070288>
- Tuerlinckx, F., De Boeck, P., & Lens, W. (2002). Measuring needs with the Thematic Apperception Test: A psychometric study. *Journal of Personality and Social Psychology*, 82, 448–461. <https://doi.org/10.1037//0022-3514.82.3.448>
- Verguts, T., & De Boeck, P. (2000). A Rasch Model for Detecting Learning While Solving an Intelligence Test. *Applied Psychological Measurement*, 24, 151–162. <https://doi.org/10.1177/01466210022031589>
- Verhelst, N. D., & Glas, C. A. W. (1993). A dynamic generalization of the Rasch model. *Psychometrika*, 58, 395–414. <https://doi.org/10.1007/BF02294648>
- Wainer, H., Bradlow, E. T., & Wang, X. (2007). *Testlet response theory and its applications*. Cambridge University Press.
- Weiss, D. J. (1982). Improving Measurement Quality and Efficiency with Adaptive Testing. *Applied Psychological Measurement*, 6, 473–492. <https://doi.org/10.1177/014662168200600408>
- Winter, D. G., John, O. P., Stewart, A. J., Klohnen, E. C., & Duncan, L. E. (1998). Traits and motives: Toward an integration of two traditions in personality research. *Psychological Review*, 105, 230–250. <https://doi.org/10.1037/0033-295X.105.2.230>
- Woodhouse, B., & Jackson, P. (1977). Lower bounds for the reliability of the total score on a test composed of non-homogeneous items: II: A search procedure to locate the greatest lower bound. *Psychometrika*, 42, 579–591. <https://doi.org/10.1007/BF02295980>

Wright, B. D., & Stone, M. H. (1979). *Best Test Design*. Univ of Chicago Social Research.

Wright, R. A. (2016). Motivation theory essentials: Understanding motives and their conversion into effortful goal pursuit. *Motivation and Emotion*, 40(1), 16–21.

<https://doi.org/10.1007/s11031-015-9536-4>

Footnotes

¹ A different parametrization of these ideas was suggested by Revelle (1986).

² We use `calc.derivs=F` because in the current `lme4` version the warnings are still under development. We followed the suggestions of the `lme4` authors and refitted the model with different optimizers to check the results.

Figure 1

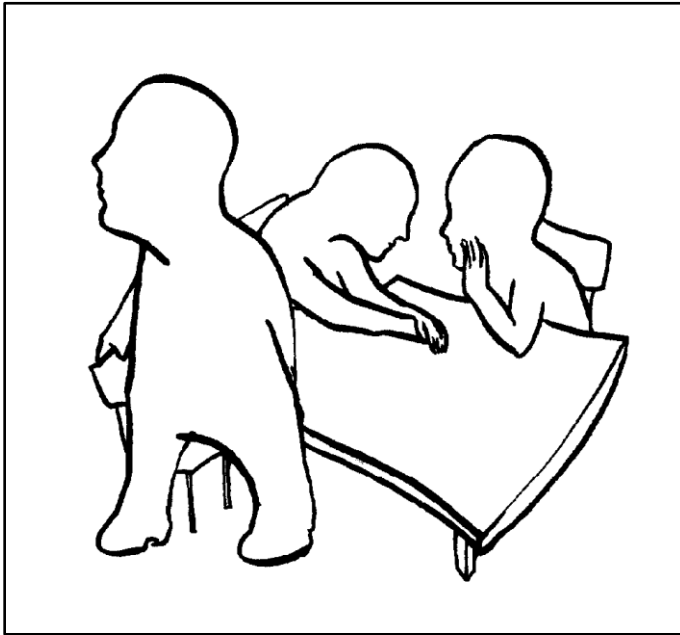


Table 1

Formulas Model M_1 - M_4

| Latent utility | Effects added by model | | | | |
|----------------|------------------------|-------|---------------------------------------|-----|--|
| | M_1 | M_2 | M_3 | or | M_4 |
| μ_{aff} | $= v_{affs}$ | $+$ | $\sum_{m=1}^M x_{affms} \beta_{affm}$ | $+$ | $TDE_{affs} \beta_{affm}$ or $SDE_{affs} \beta_{affm}$ |
| μ_{ach} | $= v_{achs}$ | $+$ | $\sum_{m=1}^M x_{achms} \beta_{achm}$ | $+$ | $TDE_{achs} \beta_{achm}$ or $SDE_{achs} \beta_{achm}$ |
| μ_{pow} | $= v_{pows}$ | $+$ | $\sum_{m=1}^M x_{powms} \beta_{powm}$ | $+$ | $TDE_{pows} \beta_{powm}$ or $SDE_{pows} \beta_{powm}$ |
| μ_{zero} | $= 0$ | | | | |

Table 2
Descriptive Information

| Picture | Affiliation % | Achievement % | Power % | Other % |
|---------|---------------|---------------|---------|---------|
| 1 | 31.4 | 0.0 | 36.9 | 31.7 |
| 2 | 70.5 | 0.2 | 18.4 | 10.9 |
| 3 | 18.1 | 3.6 | 13.3 | 65.0 |
| 4 | 6.3 | 0.3 | 56.0 | 37.4 |
| 5 | 4.2 | 9.6 | 1.4 | 84.8 |
| 6 | 0 | 65.0 | 0.6 | 34.4 |
| 7 | 0.9 | 23.1 | 41.9 | 34.1 |
| 8 | 0.5 | 2.4 | 38.0 | 59.1 |
| 9 | 2.7 | 43.0 | 6.6 | 47.7 |
| 10 | 1.4 | 3.0 | 40.2 | 55.4 |
| 11 | 3.5 | 6.0 | 72.8 | 17.7 |
| 12 | 5.8 | 15.1 | 66.7 | 12.4 |
| 13 | 10.2 | 5.7 | 40.3 | 43.8 |
| 14 | 16.2 | 5.2 | 39.7 | 38.9 |
| 15 | 0.5 | 0.5 | 39.6 | 59.4 |

Table 3

Model Comparisons

| Thurstonian Model | <i>df</i> | <i>logLik</i> | Deviance | AIC | BIC | Δdf | χ^2 |
|----------------------------------|-----------|---------------|----------|--------|--------|-------------|----------|
| M_1 : Motivational Conflict | 7 | -13,942.9 | 27,886 | 27,900 | 27,956 | | |
| M_2 : Picture effects | 42 | -10,020.1 | 20,040 | 20,124 | 20,461 | | |
| M_1 vs. M_2 | | | | | | 35 | 7,846* |
| M_3 : Picture effects with TDE | 45 | -9,985.2 | 19,970 | 20,060 | 20,422 | | |
| M_2 vs. M_3 | | | | | | 3 | 69.85* |
| M_4 : Picture effects with SDE | 45 | -9,602.2 | 19,204 | 19,294 | 19,656 | | |
| M_2 vs. M_4 | | | | | | 3 | 835.9* |

Note. TDE = temporary dynamic effect. SDE = sustained dynamic effect $k = 22,601$ pair-wise comparisons nested in $n = 633$ persons and 15 pictures. * $p < .0001$.

Table 4

Squared Correlation Reliability ($r_{\theta\hat{\theta}}^2$) for M_4 and Traditional Measures of Reliability in Bias-Corrected Simulated Data

| Motive | $r_{\theta\hat{\theta}}^2$ | Cronbach's α | glb |
|-------------|----------------------------|---------------------|-----|
| Affiliation | .52 | .17 | .29 |
| Achievement | .62 | .28 | .40 |
| Power | .73 | .46 | .56 |

$r_{\theta\hat{\theta}}^2$ = squared correlation reliability estimated by squaring the correlation between the true underlying latent motives θ and the estimated maximum a-posteriori estimates of the latent motives $\hat{\theta}$ from M_3 . glb = greatest lower bound estimate of reliability (Woodhouse & Jackson, 1977) estimated using the glb.algebraic function in the psych package (Revelle, 2016; version 1.6.9) in the R environment (R Core Team, 2014).

Table 5

Parameter Estimates for Model M4

| Estimate | M_4 | | | Average across 100 simulated datasets based on M_4 | | | Average across 100 simulated datasets based on M_4 and bias-corrected | | |
|---------------------------|-------|-------|-------|--|-------|-------|---|-------|-------|
| | Aff | Ach | Pow | Aff | Ach | Pow | Aff | Ach | Pow |
| Fixed effects (β) | | | | | | | | | |
| Picture 1 | -0.38 | - | 0.05 | -0.77 | - | 0.00 | -0.36 | - | 0.01 |
| Picture 2 | 2.07 | - | 0.42 | 2.17 | - | 0.35 | 2.31 | - | 0.44 |
| Picture 3 | 0.01 | -2.84 | -1.02 | 0.79 | -3.25 | -0.79 | 0.39 | -2.87 | -1.00 |
| Picture 4 | -0.03 | - | 0.57 | 0.76 | - | 0.55 | 0.22 | - | 0.56 |
| Picture 5 | -0.72 | -2.15 | -2.22 | 0.40 | -2.24 | -1.70 | -0.41 | -2.14 | - |
| Picture 6 | - | 0.77 | - | - | 0.76 | - | - | 0.81 | - |
| Picture 7 | - | 0.17 | 0.60 | - | 0.38 | 0.67 | - | 0.27 | 0.60 |
| Picture 8 | - | -1.53 | 0.25 | - | -1.00 | 0.48 | - | -1.45 | 0.29 |
| Picture 9 | -0.51 | 0.85 | -0.66 | 0.45 | 1.30 | -0.30 | -0.27 | 1.00 | -0.61 |
| Picture 10 | -0.94 | -0.81 | 0.51 | 0.15 | 0.00 | 0.77 | -0.73 | -0.64 | 0.54 |
| Picture 11 | 0.52 | 0.51 | 1.90 | 1.46 | 1.14 | 1.90 | 0.72 | 0.67 | 1.88 |
| Picture 12 | 1.11 | 1.41 | 2.30 | 2.04 | 2.04 | 2.34 | 1.39 | 1.60 | 2.33 |
| Picture 13 | 0.77 | 0.01 | 1.30 | 1.94 | 0.87 | 1.65 | 1.09 | 0.21 | 1.38 |
| Picture 14 | 1.33 | 0.14 | 1.50 | 2.61 | 1.02 | 1.87 | 1.65 | 0.35 | 1.58 |
| Picture 15 | - | - | 1.32 | - | - | 1.82 | - | - | 1.45 |
| SDE | -1.34 | -0.91 | -0.34 | -2.16 | -1.38 | -0.43 | -1.55 | -1.03 | -0.36 |
| Random effects | | | | | | | | | |
| SD | 1.66 | 1.49 | 0.98 | 2.43 | 1.92 | 1.02 | 1.80 | 1.58 | 0.99 |
| Correlations | | | | | | | | | |
| Aff | - | | | - | | | - | | |
| Ach | 0.19 | - | | 0.11 | - | | 0.09 | - | |
| Pow | 0.02 | 0.31 | - | -0.10 | 0.20 | - | -0.08 | 0.26 | - |

Note. $k = 22,601$ pair-wise comparisons nested in $n = 633$ persons and 15 pictures. Aff = affiliation vs other. Ach = achievement vs other. Pow = power vs other. SDE = sustained dynamic effect operationalized as the number of previous motive related responses.

Table 6

Recoded OMT data matrix

| Person | Picture | Code | DAFF | DACH | DPOW | TDAFF | TDACH | TDPOW | SDAFF | SDACH | SDPOW | Response1 | Response2 |
|--------|---------|------|------|------|------|-------|-------|-------|-------|-------|-------|-----------|-----------|
| 1 | 1 | AFF | 1 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | AFF | 1 | 0 | -1 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 1 | AFF | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 |
| 1 | 2 | ACH | 1 | -1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 1 |
| 1 | 2 | ACH | 0 | 1 | -1 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 1 | 2 | ACH | 0 | 1 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | 0 |
| 1 | 3 | POW | 1 | 0 | -1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 3 | POW | 0 | 1 | -1 | 0 | 1 | 0 | 1 | 1 | 0 | 0 | 1 |
| 1 | 3 | POW | 0 | 0 | 1 | 0 | 1 | 0 | 1 | 1 | 0 | 1 | 0 |
| 1 | 4 | ZERO | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 4 | ZERO | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |
| 1 | 4 | ZERO | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 |

Note. DAFF, DACH, and DPOW are dummy variables to indicate the three comparisons for each coding. TDAFF, TDACH and TDPOW indicate the answer to the previous picture for the respective motive. SDAFF, SDACH, and SDPOW indicate all previous answers for the respective motive. Response 1 and Response 2 are binary coded response vectors.

Appendix A

```

library(mvtnorm)

tv<-c(
  picturenAff=c(-0.38,2.07,0.01,0.03,-0.72,NA,NA,NA,-0.51,-0.94,0.52,1.11,0.77,1.33,NA),
  picturenAch=c(NA,NA,-2.84,NA,-2.15,0.77,0.17,-1.53,0.85,-0.81,0.51,1.41,0.01,0.14,NA),
  picturenPow=c(0.05,0.42,-1.02,0.57,-2.22,NA,0.60,0.25,-0.66,0.51,1.90,2.30,1.30,1.50,1.32),
  de=c(-1.34,-0.91,-0.34),
  ranefsd=c(1.66,1.49,0.98),
  ranefcor=c(0.19,0.02,0.31),
  n=633
)

genomtdat<-function(tv){
  rp=tv[paste("ranefcor",1:3,sep="")]
  sp=tv[paste("ranefsd",1:3,sep="")]
  res <- rmvnorm(tv[c("n")], mean = rep(0,3),
    sigma = matrix(c(1,
      rp[1],rp[2],rp[1],1,rp[3],rp[2],rp[3],1),3,byrow=T)*sp%*t(sp))

  pes=matrix(tv[1:45],nrow=15,ncol=3)
  idata<-matrix(c(rep(1:tv[c("n")],each=15*6),rep(as.vector(res),each=15*6),

    rep(rep(1:15,each=6),tv[c("n")]),rep(NA,tv[c("n")]*15*6*12)),ncol=17,nrow=tv[c("n")]*15*6
  )

  for (i in 1:tv[c("n")]) {
    for (j in 1:15) {

      if (j==1) { tonext<-c(0,0,0) }
      if (j==1) { tonext2<-c(0,0,0) }
      uvnr<-c(res[i,]+pes[j,]+tonext*tv[46:48],0)

      index=(i*15*6-(15*6-1))+(j-1)*6
      a<-ifelse(is.na(uvnr),-1e5,uvnr)
      ra=rnorm(4,a,1)

      aff=ifelse(ra[1]==max(ra),1,0)
      ach=ifelse(ra[2]==max(ra),1,0)
      pow=ifelse(ra[3]==max(ra),1,0)
      rest=ifelse(ra[4]==max(ra),1,0)

      idata[index,6:17]<-c(ifelse(is.na(uvnr[1]),NA,aff),ifelse(is.na(uvnr[2]),NA,ach),1,-
        1,0,0,tonext,tonext2)
      idata[index+1,6:17]<-c(ifelse(is.na(uvnr[1]),NA,aff),ifelse(is.na(uvnr[3]),NA,pow),1,0,-
        1,0,tonext,tonext2)
      idata[index+2,6:17]<-c(ifelse(is.na(uvnr[2]),NA,ach),ifelse(is.na(uvnr[3]),NA,pow),0,1,-
        1,0,tonext,tonext2)
      idata[index+3,6:17]<-c(ifelse(is.na(uvnr[1]),NA,aff),rest,1,0,0,-1,tonext,tonext2)
      idata[index+4,6:17]<-c(ifelse(is.na(uvnr[2]),NA,ach),rest,0,1,0,-1,tonext,tonext2)
      idata[index+5,6:17]<-c(ifelse(is.na(uvnr[3]),NA,pow),rest,0,0,1,-1,tonext,tonext2)

      tonext<-c(ifelse(aff>0,tonext[1]+1,tonext[1]),ifelse(ach>0,tonext[2]+1,tonext[2]),
        ifelse(pow>0,tonext[3]+1,tonext[3]))
      tonext2<-c(ifelse(aff>0,1,0),ifelse(ach>0,1,0),
        ifelse(pow>0,1,0))
    }
  }
  idata<-data.frame(idata[idata[,6]!=0|idata[,7]!=0,])
  idata<-na.exclude(idata)
  names(idata)<-c("person","trueaff","trueach","truepow","picture",

    "response1","response2","DAFF","DACH","DPOW","d4","SDAFF","SDACH","SDPOW","TDAFF","TDACH",
    "TDPOW")
  return(idata)
}
tdata<-genomtdat(tv)

```